RESEARCH NOTE

Questioned validity of Gene Expression Dysregulated Domains in Down's Syndrome [version 1; referees: 2 approved]

Long H. Do, William C. Mobley, Nishant Singhal

Department of Neurosciences, University of California, San Diego, CA, 92093, USA

Abstract

Recently, in studies examining fibroblasts obtained from the tissues of one set of monozygotic twins (i.e. fetuses derived from the same egg) discordant for trisomy 21 (Down syndrome; DS), Letourneau *et al.,* reported the presence of a defined pattern of dysregulation within specific genomic domains they referred to as Gene Expression Dysregulated Domains (GEDDs). GEDDs were described as alternating segments of increased or decreased gene expression affecting all chromosomes. Strikingly, GEDDs in fibroblasts were largely conserved in induced pluripotent cells (iPSCs) generated from the twin's fibroblasts as well as in fibroblasts from the Ts65Dn mouse model of DS. Our recent analysis failed to find GEDDs. We reexamined the human iPSCs RNAseq data from Letourneau *et al.*, and data from this same research group published earlier examining iPSCs from the same monozygotic twins. An independent analysis of RNAseq data from Ts65Dn fibroblasts also failed to confirm presence of GEDDs. Our analysis questions the validity of GEDDs in DS.

Open Peer Review

Referee Status: ☑☑

	Invited Referees	
	1	**2**
version 1 published 17 Jul 2015	☑ report	☑ report

1   **Jon Pierce-Shimomura**, University of Texas at Austin USA, **Sarah Nordquist**, University of Texas at Austin USA

2   **Roger Reeves**, John Hopkins University USA

Discuss this article

Comments (0)

Associated Editorial

Do L and Mobley W » Single Figure Publications: Towards a novel alternative format for scholarly communication, *F1000Research* 2015, **4**:268 (doi: 10.12688/f1000research.6742.1)

Corresponding author: Nishant Singhal (nisinghal@ucsd.edu)

How to cite this article: Do LH, Mobley WC and Singhal N. **Questioned validity of Gene Expression Dysregulated Domains in Down's Syndrome [version 1; referees: 2 approved]** *F1000Research* 2015, **4**:269 (doi: 10.12688/f1000research.6735.1)

Competing interests: The authors declare no competing financial interests.

First published: 17 Jul 2015, **4**:269 (doi: 10.12688/f1000research.6735.1)

Main text

The surprising fndings by Letourneau[1] and colleagues prompted us to examine our own, as yet unpublished, Ts65Dn transcriptome data for the developing and mature hippocampus in an attempt to identify GEDDs. Our data provided no evidence for the pattern reported in Letourneau *et al.*'s work. We first entertained the possibility that GEDDs were not present in post mitotic cells or cells undergoing neural differentiation. However, to ensure that we fully understood the published GEDD data, we examined the entire RNAseq dataset from the Letourneau manuscript, as provided publicly via the Gene Expression Omnibus.

Principal component analysis (PCA) of RNAseq replicates from the twin's fibroblast (T1DS: twin with DS; T2N: disomic twin) revealed a great deal of variability (Figure 1A). When the datasets from Letourneau *et al.*, are compared, in two of four cases, a closer relationship exists between the DS and disomic twin fibroblasts than for replicates from the same individual. [The datasets from Letourneau *et al.* are denoted by –L]. For example, one of the 2N-hFibro-L replicates clustered more tightly with a DS-hFibro-L replicate than with any of its own replicate (2N) samples.

We next checked the variability of the twin's iPSCs RNAseq data. We found an additional three RNASeq replicates (hiPSCs-H) performed by the same research group and published earlier using fibroblasts from the same monozygotic twins[2]. PCA analysis of these data revealed that replicates of hiPSCs-H (H for Hibaoui) clustered well together; however, they did not cluster well with the data for hiPSCs-L (Figure 1A). Altogether, PCA analysis indicated marked variability between datasets, raising the possibility that technical issues in the RNAseq samples or in their analysis compromised the Letourneau study.
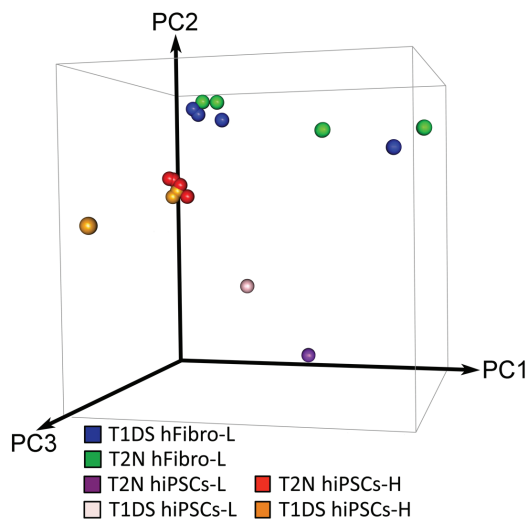
To further explore the additional three hiPSCs-H RNAseq replicates, we searched for GEDDs using methods similar to those utilized by Letourneau and colleagues. Our analysis of the hiPSCs-H did not find conserved patterns indicative of GEDDs. Figure 1B shows the results for two chromosomes, as examples. The authors reported high global gene fold-change correlations between the twin's fibroblasts and derived iPSCs. Our re-analysis found a similar high correlation value of $\rho = 0.82$ between the hiPSCs-L and hFibro-L. However, we found the hiPSCs-H poorly correlated with the original datasets; $\rho = 0.31$ between hiPSCs-L and hiP-SCs-H; $\rho = 0.07$ between hiPSCs-H and hFibro-L (Figure 1C and Supplementary figure 1A).

Conservation of GEDDs in Ts65Dn mouse model of DS were quite unexpected given that Ts65Dn mouse is segmentally trisomic (34Mb) for a portion of mouse chromosome 16 (MMU16); the segment contains about 88 mouse homologues to human genes on the
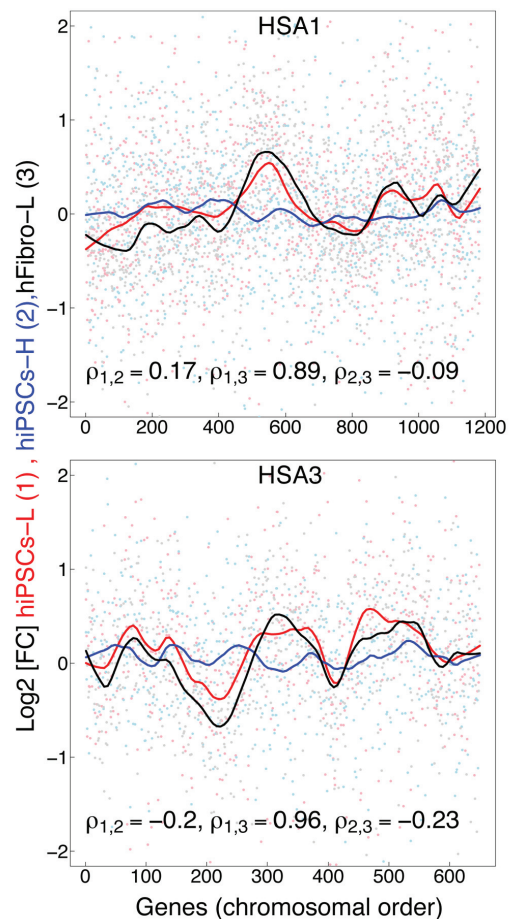

Figure 1B. Comparison of the gene expression fold-change profiles between T1DS and T2N in human fibroblasts and human iPS cells along human chromosomes, HSA1 and HSA3. (1) hiPSCs-L derived from monozygotic twins discordant for DS, Letourneau 2014 (red). (2) hiPSCs-H from the same monozygotic twins discordant for DS from Hibaoui 2014 (blue). (3) Human fibroblasts (hFibro-L) from monozygotic twins discordant for DS from Letourneau 2014 (black). hiPSCs-H lack GEDDs, while original hiPSCs-L and hFibro-L show GEDDs and high Spearman's correlation ($\rho_{1,3}$).

Figure 1A. Principal Component Analysis (PCA) of global gene expression (RNAseq) replicates from monozygotic twins discordant for DS. PCA analysis of global gene expression among RNAseq replicates from the twin's fibroblasts and iPSCs. Comparing hFibro-L replicates with themselves reveals a high degree of variability along the most significant component, PC1. In addition, there is great variability between hiPSCs-L and hiPSCs-H.

Chr (HSA)	ρ (hFibro-L, hiPSCs-L)	ρ (hFibro-L, hiPSCs-H)	ρ (hiPSCs-L, hiPSCs-H)
1	0.89	-0.09	0.17
2	0.85	-0.33	0.12
3	0.96	-0.23	-0.2
4	0.49	0.11	0.83
5	0.87	0.08	0.27
6	0.87	0.73	0.91
7	0.84	-0.11	-0.33
8	0.91	-0.77	-0.63
9	0.97	0.71	0.7
10	0.82	0	0.22
11	0.88	-0.66	-0.71
12	0.93	-0.11	-0.18
13	0.41	0.28	0.52
14	0.96	0.84	0.8
15	0.91	0.62	0.78
16	0.82	-0.22	0.03
17	0.83	0.02	0.25
18	0.71	0.1	0.6
19	0.53	0.46	0.24
20	0.4	-0.02	0.82
21	0.81	-0.1	-0.18
22	0.61	0.7	0.24
X	0.98	0.35	0.39
ALL	0.82	0.07	0.31

Figure 1C. **Correlation of global gene expression fold-change profiles from iPSCs and fibroblasts along all human chromosomes.** While iPSCs (hiPSCs-L) from the original study are highly similar to fibroblasts (hFibro-L) with a global correlation of ρ=0.82, hiPSCs-H show poor correlation with those samples (ρ=.07, ρ=0.31).

long arm of HSA21; it also carries an extra copy of the approximately 10Mb centromeric segment of MMU17 that is not syntenic to any region on human chromosome 21[3–6]. To further explore the possibility of GEDDs, RNAseq data was obtained from three replicates each from Ts65Dn and wild type mouse embryonic fibroblasts (MEFs-D) (D for Do denotes MEFs in the current study). PCA revealed tight clustering between our replicates (Figure 1D), but not those for the MEFs-L samples. While we found expected changes in gene expression in MEFs-D, analysis of MEFs-L and MEFs-D found a poor global correlation (ρ = -0.31) (Figure 1E and Supplementary figure 1B); this was also the case across all mouse chromosomes (for examples see Figure 1F and Supplementary figure 1B). In summary our findings raise serious concerns regarding the validity of GEDDs. We find no evidence for such domains in the studies on DS referenced herein or in cells from the Ts65Dn mouse model of DS.

Methods

Total RNA was collected using TRIzol reagent and further purified using RNeasy mini Kit, (Qiagen) from primary mouse embryonic fibroblasts (MEFs) derived from 18.5-day-old Ts65Dn and 2N mouse using manufacturer's instructions. RNA quality was checked using Tapestation 2200 (Agilent technologies) and quantified using
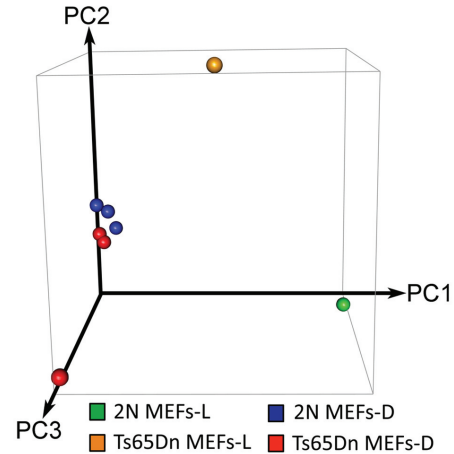

Figure 1D. **PCA of global gene expression (RNAseq) from normal (2N) and DS model mice (Ts65Dn) embryonic fibroblasts.** PCA reveals little variance along the most significant component, PC1, of global gene expression among RNAseq replicates from our repeated experiments, MEFs-D. RNAseq data of mice fibroblasts from the original study, MEFs-L, do not cluster with MEFs-D.
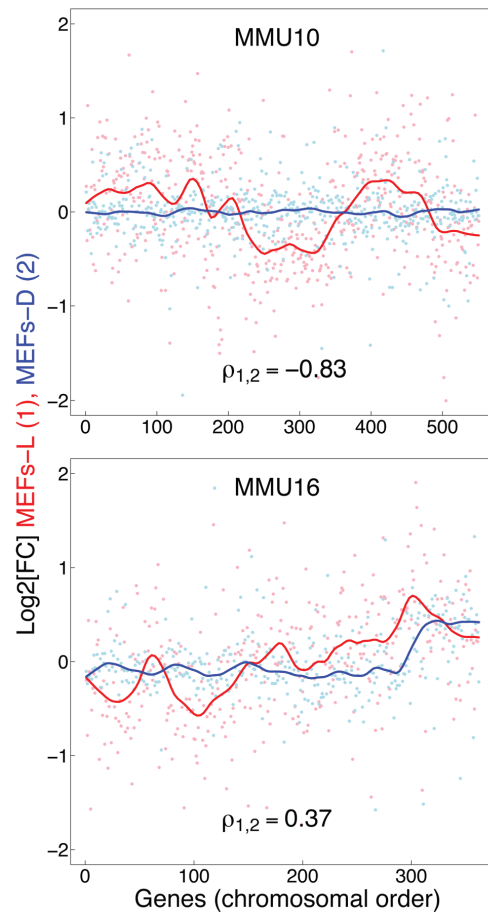

Figure 1E. **Comparison of the gene expression fold-change profiles between 2N and Ts65Dn fibroblasts plotted with respect to mouse chromosomes 10 and 16.** Embryonic mice fibroblasts examined herein (MEFs-D; blue), from 2N and Ts65DN mice do not show the GEDDs reported for MEFs-L (red).

Chr (MMU)	ρ (MEFs-L, MEFs-D)
1	-0.29
2	-0.11
3	-0.67
4	-0.62
5	-0.8
6	0.2
7	-0.1
8	-0.4
9	-0.74
10	-0.83
11	-0.32
12	-0.24
13	-0.59
14	-0.03
15	-0.32
16	0.37
17	-0.82
18	-0.32
19	-0.32
X	-0.1
ALL	-0.31

Figure 1F. Correlation of global gene expression fold-change profiles from fibroblasts for all mouse chromosomes. Global gene expression fold-change from MEFs-D and MEFs-L show a poor Spearman's correlation overall (ρ=-0.31) as well as for each of the chromosomes.

Qubit instrument (Life technologies). TrueSeq stranded mRNA-seq libraries were prepared from 5 µg of total RNA (Illumina mRNA-seq kit, RS-122-2103) and sequenced using Illumina HiSeq 2500 PE-100 (sequences publically available from GEO, accession number: GSE64840). Experiments were performed in triplicate.

RNAseq data from hFibro-L, hiPSCs-L, and hiPSCs-H were downloaded from the Sequence Read Archive (SRP039348, SRP032928) and uploaded to Illumina BaseSpace for mapping (BaseSpace App v1.0, TopHat v2) and differential gene analysis (BaseSpace App v1.1, CuffLinks v2.1.1). PCA was performed using R (v3.1.0) from normalized gene count values (FPKM). Overall Spearman correlation values were calculated from locally weighted scatterplot smoothing (LOWESS) with 30% bandwidth between log2 (FC) gene expression of comparison samples, ordered by genes along each chromosome and plotted using R and custom scripts.

Software availability

Software access
Custom scripts for R used to calculate Spearman correlation values are available at https://github.com/lhdo/GEDDplot

Source code as at the time of publication
https://github.com/F1000Research/GEDDplot/releases/tag/V1

Archived source code as at the time of publication
http://dx.doi.org/10.5281/zenodo.19232

Software license
The MIT license

Supplementary materials

Supplemental figure S1. A) Comparison of the gene expression fold-change profiles between T1DS and T2N in human fibroblasts and human iPS cells along all human chromosomes. (1) hiPSCs-L derived from monozygotic twins discordant for DS, Letourneau 2014 (red). (2) hiPSCs-H from the same monozygotic twins discordant for DS from Hibaoui 2014 (blue). (3) Human fibroblasts (hFibro-L) from monozygotic twins discordant for DS from Letourneau 2014 (black). hiPSCs-H lack GEDDs, while original hiPSCs-L and hFibro-L show GEDDs and high Spearman's correlation ($\rho_{1,3}$). **B) Comparison of the gene expression fold-change profiles between 2N and Ts65Dn fibroblasts along all mouse chromosomes.** Embryonic mice fibroblasts MEFs-D (blue), from 2N and Ts65DN mice do not show the GEDDs reported for MEFs-L (red).

Click here to access the data.

References

1. Letourneau A, Santoni FA, Bonilla X, *et al.*: **Domains of genome-wide gene expression dysregulation in Down's syndrome.** *Nature.* 2014; **508**(7496): 345–350.
   PubMed Abstract | **Publisher Full Text**

2. Hibaoui Y, Grad I, Letourneau A, *et al.*: **Modelling and rescuing neurodevelopmental defect of Down syndrome using induced pluripotent stem cells from monozygotic twins discordant for trisomy 21.** *EMBO Mol Med.* 2014; **6**(2): 259–277.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Busciglio J, Capone G, O'Bryan J, *et al.*: **Down syndrome: genes, model systems, and progress towards pharmacotherapies and clinical trials for cognitive deficits.** *Cytogenet Genome Res.* 2013; **141**(4): 260–271.
   PubMed Abstract | **Publisher Full Text**

4. Davisson MT, Schmidt C, Reeves RH, *et al.*: **Segmental trisomy as a mouse model for Down syndrome.** *Prog Clin Biol Res.* 1993; **384**: 117–133.
   PubMed Abstract

5. Reinholdt LG, Ding Y, Gilbert GJ, *et al.*: **Molecular characterization of the translocation breakpoints in the Down syndrome mouse model Ts65Dn.** *Mamm Genome.* 2011; **22**(11–12): 685–691.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. Akeson EC, Lambert JP, Narayanswami S, *et al.*: **Ts65Dn -- localization of the translocation breakpoint and trisomic gene content in a mouse model for Down syndrome.** *Cytogenet Cell Genet.* 2001; **93**(3–4): 270–276.
   PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Referee Status: ☑ ☑

☑ **Roger Reeves**
Department of Physiology, School of Medicine, John Hopkins University, Baltimore, MD, USA

Widespread mis-regulation of the expression of disomic genes in trisomic genomes has been established for well over a decade (and seems a self-evident outcome of a trisomy such as that of Hsa21 which includes eight transcription factors, 29 microRNAs and a large number of lncRNAs among other key regulators). Recently, a pattern of chromosome regions showing up- or down- regulation of transcription in the trisomic genome based on RNAseq and substantiated with correlated regional changes in chromosome architecture was reported by Letourneau et al. (Letourneau et al., 2014). They referred to these regions as "gene expression dysregulated domains" or GEDDs. While the original analysis was carried out in fibroblast and iPS lines derived from a co-isogenic pair of monozygotic fetuses, one of which was euploid and the other of which had trisomy 21, Letourneau et al. report that the pattern of GEDDs is conserved in fibroblasts from the *outbred* Ts65Dn trisomic mouse model of Down syndrome (DS), i.e., the same blocks of disomic genes are up- or down-regulated in a mouse trisomic for orthologs of about half of the genes on Hsa21. These same effects are reported to be undetectable in a pooled comparison of 8 trisomic and 8 euploid cell lines, where they are hypothesized to be masked by normal variability in gene expression expected in individuals who are (obviously) not co-isogenic.

In this report, Do et al. were unable to reproduce these GEDD patterns in an analysis of Ts65Dn mice similar to that which formed a part of the Letourneau evidence. Do et al. then proceeded to compare the Letourneau RNAseq results in this paper to RNAseq results generated previously by this group from cell lines of this same twin pair (Hibaoui et al., 2014) and found substantial variability in the results obtained in these independently done experiments. Based on several such comparisons in addition to the absence of GEDDS in their independent experiments, Do et al. conclude that there is not clear evidence for the existence of GEDDs.

The comments by Pierce-Shimomura and Nordquist provide an excellent summary of several statistical issues in the analyses. They suggest possible clarifications to the Do study and conclude, with Do, that the existence of GEDDs is not proven. I would note only two additional two points.

1. The PCA analysis in Do et al. comparing datasets from Letourneau and Hibaoui appears to show very strong batch effects, which is in fact the expected outcome for this type of comparison of two studies on different RNA sets prepared and run at different times, probably on different sequencers with different lots of reagents, without reference to placement in sequencing cells, etc. Statistical methods have been developed to clean data for this type of comparison (e.g., Leek JT, Johnson WE, Parker HS, Fertig EJ, Jaffe AE and Storey JD. *sva: Surrogate Variable Analysis*. R package version 3.16.0.). It is not clear that this has been done in the Do analysis.

2. As in all analyses of large platform datasets, it is impossible to determine from the published and supplemental methods what has actually been done at a level of detail required to reproduce it independently – this applies to both studies. (For the exception proving this rule, consider Gilad Y and Mizrahi-Man O. *A reanalysis of mouse ENCODE comparative gene expression data* [v1; ref status: indexed, http://f1000r.es/5ez] *F1000Research* 2015, **4**:121 (doi: 10.12688/f1000research.6536.1)).

However, it is immediately obvious that Letourneau et al. have no basis for many of the statistical comparisons used to identify GEDDs – a *biological* phenomenon - as they have no biological replicates, only technical ones. It was not clear to this reviewer whether Letourneau established four fibroblast clones and/or iPS lines from each twin (and if so, in how many independent transformation experiments) but this would only be a technical replicate for the artifact of that transformation process, not for the biology of GEDDs in trisomy, and therefore it is not the appropriate basis for the statistically-based conclusions that they make about the *biology* of the effects of trisomy 21 on transcription. For biological conclusions, there is an N of one euploid and one trisomic individual – statistical assessments are not possible with a single comparison.

Letourneau et al. argue that genetic variability affecting gene expression levels does not allow detection of GEDDs in any but co-isogenic conditions. However, if GEDDs exist and are so highly conserved in evolution that the same genes are mis-regulated in the same way in outbred* trisomic mice, it is difficult to understand why they are not evident in any comparison of humans with two vs. three copies of Hsa21; if GEDDs only occur in vanishingly rare, coisogenic monozygotic human twin sets discordant for a specific trisomy, the phenomenon would hardly seem worthy of the attention it has received. An explanation for the conservation of this phenomenon in outbred individuals of another species but not between human pairs would strengthen the understanding of this phenomenon. An adequately described experiment comparing multiple individuals with two vs. three copies of Hsa21 including euploid and trisomic sib pairs, in addition to the statistical clarifications suggested by Pierce-Shimomura, would help to establish the existence or not of GEDDs and provide some indication of their relevance to the goal of ameliorating effects of gene dosage in DS.

*Ts65Dn mice are maintained as an advanced intercross between any of several C57BL6 and C3H strains. Thus, individual mice and their sibs are not genetically identical and are heterozygous at ~50% of all loci - a different 50 % for each individual. SNPs occur about every 3000 bp between B6 and C3H, while SNPs for a given segment of a human chromosome pair might be on the order of one per 1000 bp. It might be important to consider variability in chr21 alleles in each twin. Assuming the original conceptus was trisomic, if trisomy resulted from a meiosis I error, the trisomic line will carry three sets of Hsa21 alleles while the euploid twin will lack one of the three sets. If the trisomy resulted from a meiosis II error, the euploid twin could be isodisomic for Chr21 and carry only a single set of alleles. One could speculate about the possible impact of isodisomy on genome-wide expression patterns.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Jon Pierce-Shimomura, **Sarah Nordquist**

Department of Neuroscience, College of Natural Sciences, University of Texas at Austin, Austin, TX, USA

Do *et al*. perform statistical analysis on previously published RNAseq data from Letourneau *et al*. (2014) and their own new data to test the prominent and intriguing new hypothesis that trisomy 21 may cause up and down regulation of groups of genes associated with specific physical domains on different human chromosomes (Letourneau *et al*., 2014). These were called gene expression dysregulation domains (GEDDs). Additional evidence suggested that GEDDs may be conserved across human cell types, and surprisingly, may relate to equivalent syntenic domains in the mouse genome after analysis of the Ts65Dn mouse model of Down syndrome (DS) (Letourneau *et al*., 2014).

Here, Do *et al*., first report that replicates from the Letourneau *et al*. (2014) dataset are more variable than one would expect as determined by principle component analysis (PCA). This is an important finding because identification of the purported GEDDs and their extrapolated conservation across tissue types and species depends on minimal variation across datasets. To better understand the significance of the apparent variation in replicate datasets (Figure 1A), it would be useful if Do *et al*. could also plot the percent variance aside each dot accounted for the first principal component. The procedure for data normalization should also be explained more thoroughly in the results section as this may significantly alter the apparent variance by PCA. Do *et al*. also find that data from the Letourneau *et al*., 2014 study varies significantly from data from a previous study by the same lab that used human cells derived from the same source in (Hibaoui *et al*., 2014).

Do *et al.* then analyze how gene expression changes across physical locations on human chromosomes. They replicate the original finding by Letourneau *et al.* (2014) showing that there are domains of up and down regulated genes from human iPSCS and fibroblasts derived from the same source. However, they also find that these domains fail to correlate with data derived from the same source but published in an earlier study from the same group (Hibaoui *et al*., 2014). This finding raises significant doubt about the concept of conserved GEDDs if they cannot be replicated from tissue derived from the same individual and collected by the same research group. Do *et al*., might do well to suggest explanations for the lack of correlation including specific analysis techniques and methodologies.

Lastly, Do *et al*. find that RNAseq datasets from wild-type and the Ts65Dn mouse model published in the (Letourneau *et al*., 2014) paper show considerable variation from their new set of mouse data as determined by PCA. They also show a lack of correlation between fold-change gene expression for both datasets across chromosomes.

Together, this new analysis suggests a re-evaluation of the GEDDs concept related DS. Specific groups of physically-linked genes (domains) may indeed be up and down regulated in DS across individuals and perhaps in mouse models of DS. However, variation within and across RNAseq datasets appears to prevent defining these domains and generalizing them to other individuals and species with current methods of analysis.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.